Haiku
Human AI teaming Knowledge and Understanding for aviation safety

Deliverable N. 3.3

Human-AI Teaming Validation Framework

Authors:

Hilburn, B., Villani, A. & Reis, R.J.

Abstract:

This D3.3 report is the third deliverable of HAIKU's WP3 (Human-AI Teaming), and covers the task 3.5 efforts to develop a provisional framework for validating the project's Use Case (UC) prototypes. A Use Case (UC) validation survey was developed based on the output of the task 3 .1 review, and EASA's (2023) recent guidance on trustworthy AI. It was administered to all six of the HAIKU UCs. Results indicated that the UCs differed in both their target AI levels and in their currently perceived validation concerns. This was seen as encouraging first evidence that—as intended—each UC captures different AI roles, benefits, and aviation needs, and that the six tap into a range of HAIT issues. Finally, a preliminary mapping was made to potential validation methods, also based in part on the earlier work of task 3 .1.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Information Table

Deliverable Number	3.3
Deliverable Title	Human-AI Teaming Validation Framework
Version	1.0
Status	Final
Responsible Partner	CHPR
Contributors	Villani, A. (EMBRT) Reis, R.J. (EMBRT) All HAIKU Use Case Leaders
Reviewers	Tuvesson, V. (LFV) Pozzi, S. (DBL) Arrigoni, V. (DBL)
Contractual Date of Delivery	June 31st, 2023
Actual Date of Delivery	July 4th, 2023
Dissemination Level	Public

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Document History

Version	Date	Status	Author	Description
0.1	May 10th, 2023	Draft	B. Hilburn (CHPR)	Structure of the document and first draft
0.2	May 15th, 2023	Draft	V. Arrigoni (DBL) S. Pozzi (DBL)	Review with comments
0.3	June 10th, 2023	Draft	B. Hilburn (CHPR) A. Villani (EMBRT) R. Reis (EMBRT)	Updated version
0.4	June 15th	Draft	HAIKU UC leaders	Contribution provided
0.5	June 23rd, 2023	Draft	B. Hilburn (CHPR)	Updated version
0.6	June 26th, 2023	Draft	S. Pozzi (DBL) V. Arrigoni (DBL)	Review with minor comments
0.8	June 27th, 2023	Draft	B. Hilburn (CHPR)	Updated version
0.9	June 28th, 2023 July 3rd, 2023	Draft	V. Tuvesson (LFV) S. Pozzi (DBL) V. Arrigoni (DBL)	Review with no com. Quality check
1.0	July 4th, 2023	Draft	B. Hilburn (CHPR)	Final version

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

List of Acronyms

Acronym	Definition
AI	Artificial Intelligence
AltMOC	Alternative Means of Compliance
AMC	Acceptable Means of Compliance
ATM	Air Traffic Management
DA	Digital Assistant
EASA	European Union Aviation Safety Agency
HAIKU	Human AI teaming Knowledge and Understanding for aviation safety
HAIT	Human AI Teaming
HAT	Human Autonomy Teaming
HF	Human Factors
IA	Intelligent Assistant
IR	Implementing Rules
LACC	Levels-of-autonomy-in-cognitive-control
LACC-LOA	A matrix of LOA and LACC, for the identification of critical HAIT issues.
MbC	Management by Consent
MbE	Management by Exception
ML	Machine Learning
MOC	Means of Compliance
SA	Situation Awareness
SOAR	State of the Art Report
UC	Use Case

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

UAM	Urban Air Mobility
UTM	Unmanned Aircraft System Traffic Management
WP	Work Package

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Executive Summary

D3.3 is the third deliverable of HAIKU's Human-AI Teaming work package (WP3) and covers task 3.5 (HAIT Validation Methods). Task 3.5 aims to develop a provisional human-centred framework for validating the project's Use Case (UC) prototypes. This requires identifying the appropriate methods, metrics, and success criteria for validation.

Task 3.5 surveyed each of the six UCs to identify the major issues relevant to each UC, that warrant special focus in validation. This task started from the output of the task 3.1 review. By integrating identified HAIT Human Factors (HF) issues across references, a categorization scheme was superimposed on the classification suggested by EASA's (2023) most recent guidance for levels 1 and 2 AI. EASA guidelines were modified somewhat to remove common items, and add additional items which EASA had not explicitly considered. A UC validation survey was then iteratively developed over three versions, pretested with a trial use case, and finally administered to each of the UC development teams. The survey was completed in two sections. In the first, UCs classified their target level(s) of subtask AI according to the EASA Level 1/2 scheme. Each UC also identified what they perceived to be the most pressing / critical potential validation issues in their UC.

Results indicated that the six Use Cases range in both their target AI levels, and in their currently perceived validation concerns. This was encouraging preliminary evidence that, as hoped, the six HAIKU Use Cases each have a slightly unique profile and capture different aspects of human – AI teaming, e.g. different AI role, benefits, aviation needs. The validation survey was not intended to be either exhaustive (pre-certification would require a fuller set of validation requirements) or compulsory. Instead it was intended to help the UCs identify the most salient potential Human Factors validation focus area going forward. It is also recognized that this effort was only a 'snapshot ' in that each UC is somewhat evolving. For this reason, we will consider read ministering (perhaps a refined version of) this survey at a later stage.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Table of Contents

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Introduction

The *Human AI teaming Knowledge and Understanding for aviation safety* (HAIKU) project aims to generate knowledge on intelligent assistants, and to develop AI enabled prototypes for six aviation-related Use Cases (UCs). This report is Deliverable 3.3 (*HAIT Validation Framework*) of the HAIKU project, and describes the work carried out under the project's subtask 3.5.

1.1. HAIKU technical workflow

The PERT chart of Figure 1 conceptually places HAIKU's WP3 (*Human-AI Teaming*) within the technical work flow of the overall project. WP2 (*Human-Centric Intelligent Assistance*) and WP3 together provide the vision and conceptual foundation for the project. WP2 laid out the vision, guiding principles, reference scenarios, and intended societal impact analysis to help drive end-user and stakeholder engagement. WP3, as described in the following section, aimed to develop human factors guidance and methods for human – AI teaming, informing the definition of Intelligent Assistants concepts.

WPs 4 (*Intelligent Assistance Development*) and 5 *(Explainability in HAIT*) represent the main technological development phase of the HAIKU project. IA development in WP 4 depends on inputs from WPs 2, 3, and 6 for guidance on aspects of societal-, human – AI teaming-, and use case validation requirements, respectively. WP5 is focusing more deeply on explainability concepts in HAIT, and on identifying human performance requirements of XAI in each of the use cases.

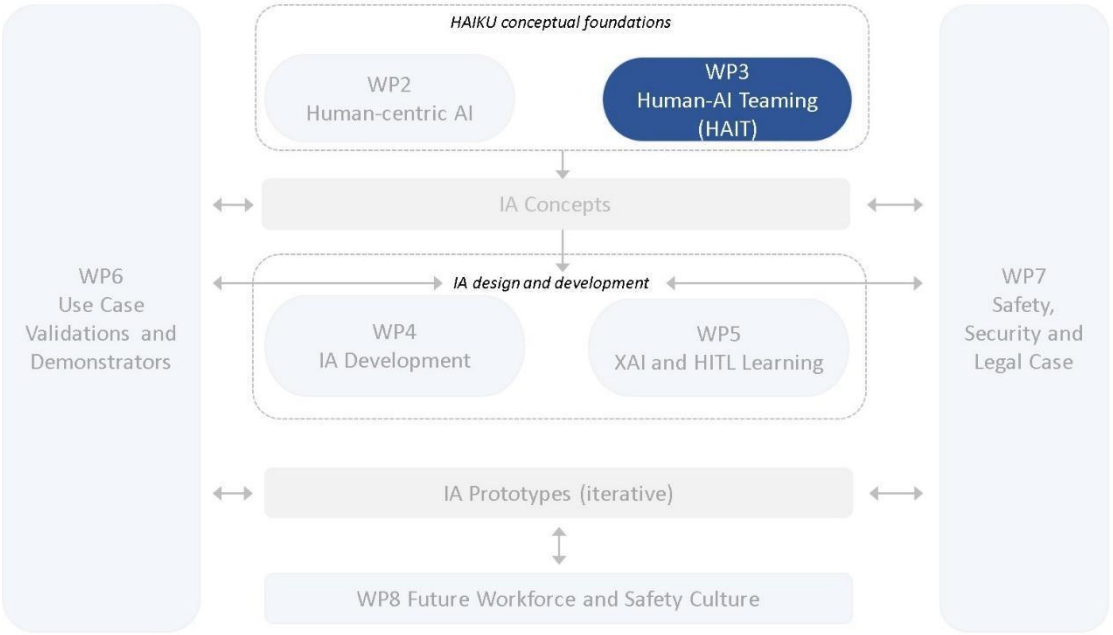


Figure 1. Work flow across the HAIKU technical WPs.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

WP6 (*Use Case Validations and Demonstrators*) interacts with the technical development WPs (WPs 4, 5 and 7) by demonstrating and conducting validation activities around the prototype IAs, separately for each of the use cases. WP7 (*Safety, Security & Legal Case for AI*) runs in parallel with HAIKU's technical development work, and aims to address safety, security, and liability issues. WP7 will perform human factors, safety, security, reliability and regulatory analyses for each use case. The final technical work package, WP8 (*Future Workforce & Safety Culture*) focuses on the implications of IA technologies across use cases, on the skills, selection, and training requirements for a future workforce.

1.2. WP3 (Human-AI Teaming)

WP3 aims to develop Human Factors design guidance and methods ('HF4AI' Capabilities) for appropriate human-AI teaming, and has the following specific objectives:

- Conduct a state-of-the art-review (SOAR) of HAIT literature;
- Refine a model and taxonomy of HAIT concepts;
- Specify the constructs underlying effective HAIT performance, and
- Develop a preliminary HAIT validation framework.

WP3 Sub tasks

As shown in the WP3 flowchart of figure 2 a total of five subtasks have been performed, resulting in three deliverables. Deliverable 3.1 (Bång et al., 2023) integrated the work of subtasks 3.1-3.2, in which the team conducted a state-of-the-art review of theoretical and empirical literature on human – AI teaming, with a focus on the aviation industry. This review identified 28 ML-based applications (from conceptual developments to prototypes to commercially available products), as well as 19 research projects, and over 100 scientific references. D3.1 also presented the LACC-LOA framework that the HAIKU project has chosen as its general approach to HAIT design. Deliverable 3.2 (Venditti, Arrigoni & Cirillo, 2023) presented the work of subtask 3.4, in which a series of four design workshops (one for each of four aviation segments: Flightdeck, ATM, UATM, and airport) produced a set of IA concepts intended to inspire HAIKU's six UCs.

Task 3.5 (HAIT Validation Methods) is the final task of WP3 (Human-AI Teaming), and the subject of this report D3.3. The overall aim of Task 3.5 was to define a provisional framework for validating the human-related aspects of the Use Case prototypes, including validation success. As described in the following sections, this involved identifying preliminary Human Factors validation concerns and issues, and linking these to potential validation criteria, methods and metrics.

It is important to note that this process was not intended to be exhaustive. We were not aiming for full scale validation, but to help UCs start thinking about the types of human issues they would have to address, and to focus attention on the unique and

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

challenging HAIT aspects around their Use Case. Further, this process was not meant to be compulsory or prescriptive for the UCs—the aim was not to 'tell them what to do.' Again, the aim was to help each UC focus on its specific concerns. Surveys of the individual Use Cases were used to help identify the main potential HAIT validation key focus aspects. These aspects are likely to evolve in step with the UCs themselves, which are currently under development and, thus, still subject to change.

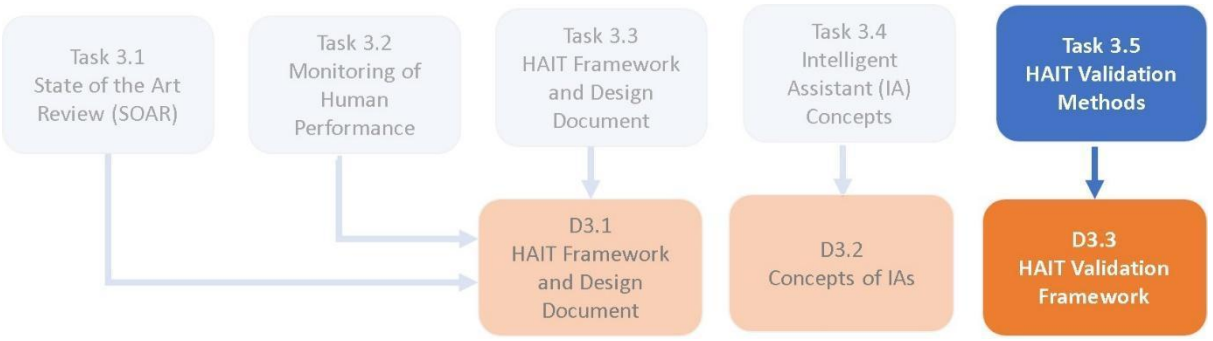


Figure 2. WP3 subtasks and deliverables.

1.3. Document structure

The remainder of this document is structured in three sections:

- Section 2 describes the methods used to develop and administer the UC validation survey;
- Section 3 presents the results of the UC survey;
- Section 4 discusses lessons learned for UC validation.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

2. Method

2.1. Inputs

Survey development was based primarily on two sources from task 3.1: the literature review (see D3.1, Bång et al, 2023) and the EASA Roadmap and Guidance documents.

The literature review identified 100+ interrelated HF constructs relevant to HAIT. These are presented in Annex D. To provide a practical (and increasingly accepted) framework for classifying potential validation issues, the recent work of EASA was also used.

EASA's AI roadmap and associated guidance.

The European Union Aviation Safety Agency (EASA) is providing guidance on the development of human – centric AI in aviation. Phase 1 of EASA's AI Roadmap project (2019-2024) has resulted in an initial Roadmap for Trustworthy AI (2020) and corresponding guidance for levels 1 and 2 AI/ML. According to EASA this guidance is a basis for their AI Roadmap, but does not provide definitive guidance on how to achieve it. Implementing rules and means of compliance (either AMCs or AltMOCs) are not yet available for AI. This guidance (in the form of objectives) is therefore presented as "an all-purpose instrument" to be customised to specific AI applications.

EASA's Roadmap (2020) is structured around the following four 'building blocks' and sub-elements for achieving "trustworthy AI and enabling readiness for use in aviation":

- Trustworthiness analysis
    - Characterization of AI
    - Safety assessment
    - Information security assessment
    - Ethics-based assessment
- AI assurance
    - Learning assurance
    - Development / post-ops explainability
- HF for AI
    - Operational explainability
    - Human AI teaming
    - Modality of interaction
- Safety risk mitigation.

EASA's Guidance document v2 (2023) classifies AI levels as follows (see also Annex B):

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

- Level 1 AL/ML (augmentation and assistance) retains full human authority, and aims to either
  - Augment human information acquisition and/or analysis processes (Level 1A); or
  - Assist human decision-making and/or action processes (Level 1B).
- Level 2 AI/ML (cooperation and collaboration) is a hybrid of full- and partial human authority, and aims to either
  - Assist decision / action selection at the sub-task level, and retain full human authority to override (Level 2A); or
  - Assist decision / action selection at the sub-task level, but the human has only partial authority to override (Level 2B).

According to this view, a Level 2A AI/ML system is 'cooperative' and works according to a predefined task allocation scheme. The AI provides the operator feedback on decision making and action implementation. A Level 2B system, on the other hand, is 'collaborative' and works with the operator to achieve a common goal. Unlike Level 2A AI, Level 2B brings requirements for shared situation awareness between human and AI, dynamic strategy adaptation, and real-time task reallocation between human and AI. According to this view, level 2B places much greater demands on communication between humans and AI. EASA's definitions of levels 2A and 2B seem to parallel the broad notions of "Management by Consent"(MbC) and "Management by Exception" (MbE) although this distinction is not always helpful (Westin et al., 2013). For example, a given system might use a time-out logic that forces the user to respond within a fixed interval, after which the system auto-implements. Such a system would have features of both MbC and MbE. To the user it looks like an MbC system, until the countdown expires.

EASA's classification scheme does not reflect a continuous Levels of Automation (LOA) scale. Over the years, various LOA scales and taxonomies have been proposed, and they have tended to define levels based on (sometimes non-orthogonal) combinations of human / machine authority, autonomy, control structure, and information processing stages. Two of the best-known LOA taxonomies appear to be those of Sheridan & Verplank (1978) and Parasuraman, Sheridan & Wickens (2000). Unlike many earlier frameworks, EASA's classification scheme seems more functionally grounded, although it seems to implicitly map Level 1 onto Information Acquisition tasks, Level 2 on Decision Making, and Level 3 on Action Implementation, thereby assuming an implicit hierarchy of tasks and related cognitive functions.

Figure 3 maps human / AI authority structure onto the customary stages of information processing (input, decision, action). According to EASA (2023) the major distinction among levels 1B-2B lies in the human/machine authority structure over decision- and action selection. Notice that EASA does not currently provide guidance on Level 3 AI, in which:

- Level 3A: AI is responsible for action selection and implementation, but the human is available 'on request'
- Level 3B: The system operates fully autonomously, with no human intervention.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

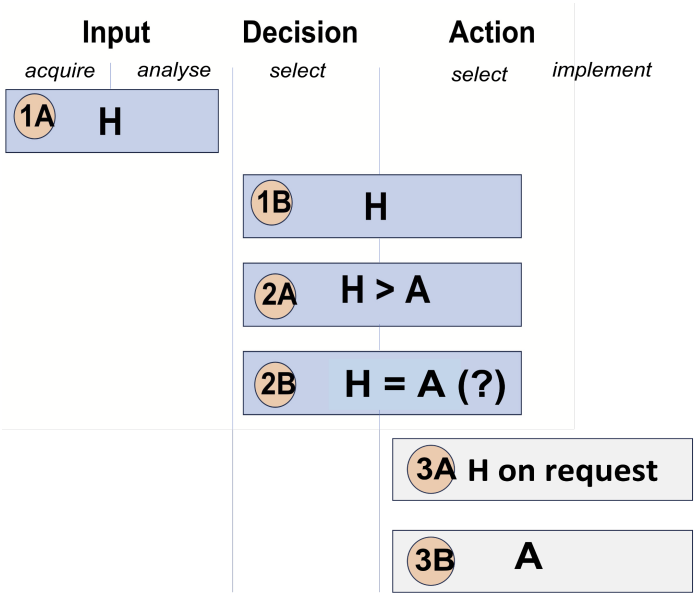


Figure 3. EASA's AI Level classification (adapted from EASA (2023)).

2.2. Survey development

The validation survey was iteratively developed over three versions, and pre-tested through application to one test UC, before administration to all six UC development teams. For completeness, EASA objectives were cross-checked against the HF constructs identified in WP3.1 review. These resulted in a hybrid instrument that included line references for EASA objective items, along with HF constructs relevant to each category of objectives. In several cases some concepts identified in the D3.1 SOAR (such as Situation Awareness) were not explicitly covered in EASA's list, and these were included. Certain elements (e.g. certain EASA "ConOps" items related to role definition) were excluded, as they were seen to be common across UCs.

Again, the main goal of the survey was to assess the relevance of HAIT constructs for each of the UCs, as judged by the UC teams. The final version of the survey is shown in Annex C, and groups HAIT constructs under the following eight categories:

- Explainability / transparency
- Situation Awareness
- Ethics
- Collaboration and teamwork
- Workload
- Information security
- Safety assessment and assurance
- Objective performance criteria

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

An excerpt of the UC survey is shown in Figure 4, for one of the eight categories. The table columns cover EASA objective, EASA reference (from EASA, 2023), AL/ML level, and corresponding HF issues identified in task 3.1.

Collaboration and teamwork

☐	Ensure two-way cross check of proposals	HF-04/05/06/11	2A-2B	☐ Acceptance
☐	Identify suboptimal performance or abnormal operation	HF-05 / HF-06	2B	☐ Autonomy
☐	Ensure bi-directional communication	HF-02	2B	☐ Complexity, perceived
☐	Ensure AI can build its own Situation Awareness	HF-01	2B	☐ Individual differences
☐	Ensure AI can modify its own Situation Awareness	HF-03	2B	☐ Reliability
☐	Notify operator of his / her misunderstanding	HF-12	2B	☐ Reliance, over / under
				☐ Self-confidence
				☐ Trust

Figure 4. Excerpt of the Use Case survey (see also Annex C).

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

3. Results

3.1. AI classification level, by UC

	UC1 Startle	UC2 Flt plan	UC3 UAM	UC4 Dig Twr	UC5 Arpt Safety	UC6 Pandemic
Level 1A Human augmentation (human has full authority)						
support to information acquisition						x
support to information analysis						x
Level 1B Human assistance (human has full authority)						
Support to decision selection	x			x	x	x
Support to action selection	x			x	x	x
Level 2A Human – AI Cooperation (human has full authority)						
Cooperative overridable automatic decision selection	x	x	x			x
Cooperative overridable automatic action selection	x		x			
Level 2B Human – AI Collaboration (human has partial authority)						
Collaborative overridable automatic decision selection		x				x
Collaborative overridable automatic action selection						

Figure 5. Target AI level(s), by Use Case.

Figure 5 shows the range of target AI levels across UCs. Xs indicate targeted levels / elements of AI, and graduated shades of red indicate highest target AI level. UCs 3, 4 and 5 target only one AI level. The other three UCs (1, 2, and 6) target more than one AI level. UC6 targets elements of all four AI levels, UC1 targets levels 1B and 2A, and UC2 levels 2A and 2B.

It was interesting to see that **there are five AI level profiles across the six UCs**. Only UCs 4 and 5 (digital tower and airport safety) share AI level profile—both are targeting level 1B automation, with full human authority of both decision- and action selection. UC2 and UC6, which target level 2B AI, both aim for fully automated **decision** selection, but stop short of fully automated **action** selection.

3.2. EASA category and item weighting

UC development teams indicated which of the EASA guidance items (except for situation awareness and workload objectives, each guidance item has a corresponding EASA reference number) were relevant to their UC, and where able the teams provided a ranking of the EASA items judged most critical. Related concepts (from D3.1) were included mainly to clarify the definition of each category, and UC teams also provided rankings of these related concepts where able.

Within each category, ranked items (n=1-x) were binary split into High and Medium criticality (abbreviated as H and M in Annex D). These ranked items were identified as 'flagged' and survey responses were then processed separately for each UC as follows: for each of the eight categories, raw category weighting was calculated as the number of flagged- versus- total category items

Raw weighting= number of flagged category items / total number of category items

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Raw proportions were then standardised within UC, such that the category weights for each UC sum to 1 (disregarding rounding errors). These results are shown in Table 1 and Figure 6.

	Explainability	SA	Ethics	Teamwork	Info Security	Safety Assmnt	Obj Perf	Workload
UC1 (Startle)	0.10	0.13	0.13	0.00	0.00	0.16	0.21	0.26
UC2 (Plan)	0.08	0.10	0.08	0.10	0.20	0.12	0.12	0.20
UC3 (UAM)	0.22	0.29	0.07	0.19	0.00	0.23	0.00	0.00
UC4 (DigTwr)	0.08	0.13	0.10	0.00	0.25	0.05	0.15	0.25
UC5 (Airport)	0.19	0.35	0.09	0.12	0.00	0.07	0.18	0.00
UC6 (Pandemic)	0.06	0.19	0.07	0.09	0.19	0.11	0.11	0.19

Table 1. EASA category weighting (standardised within UC).

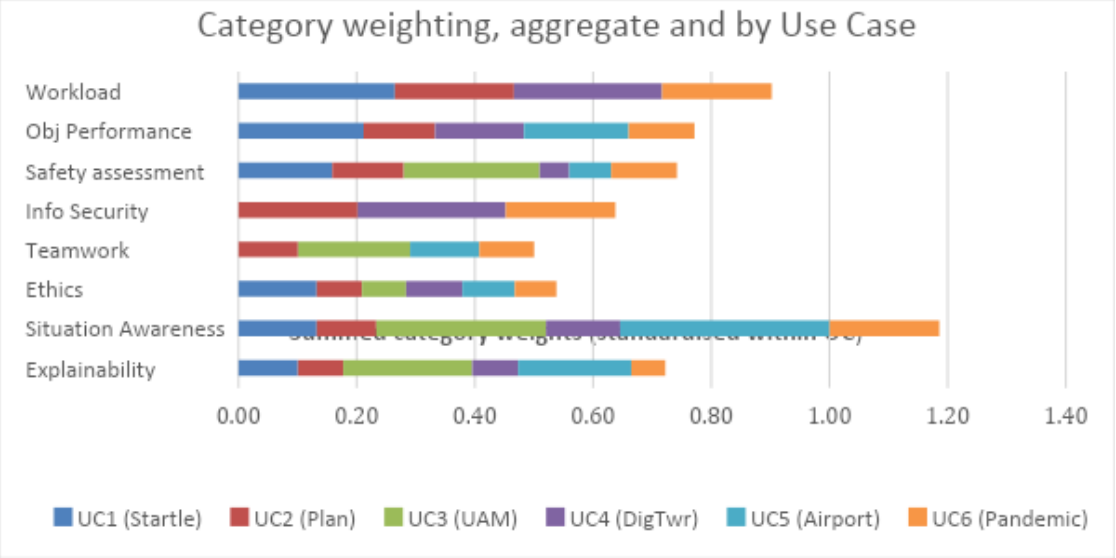


Figure 6. EASA category weighting.

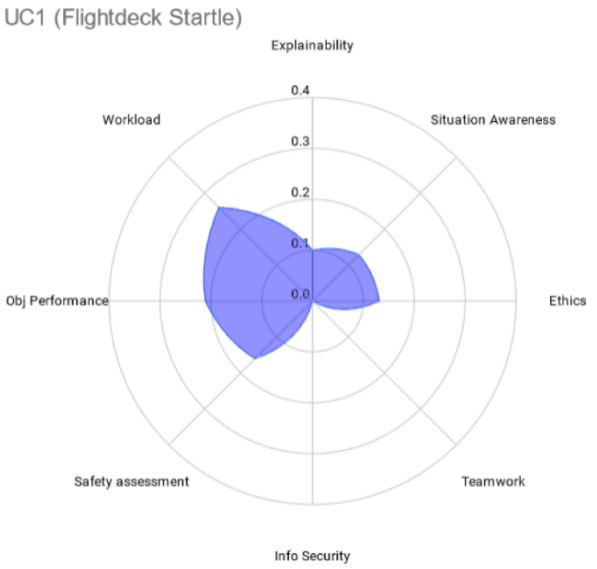
Across UCs, Situation Awareness and Workload were the two highest-rated categories (however these two categories had a small number of items, which complicates analysis). Teamwork and Ethics were the lowest-rated categories.

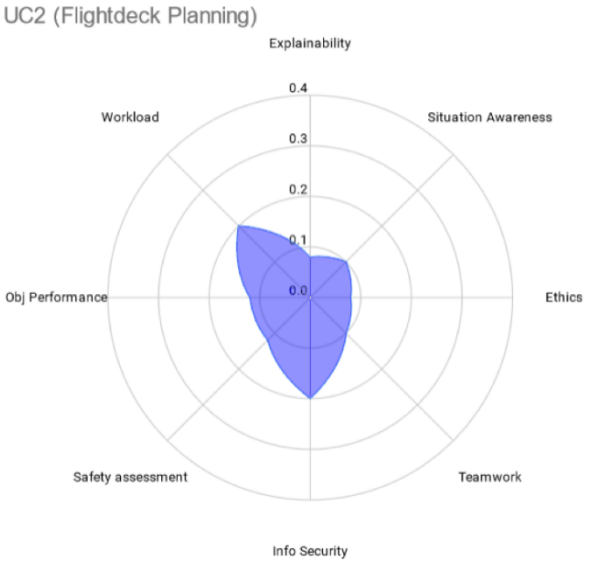
Spider graphs of the eight categories are shown in figures 7a-f. Each UC seems to have a unique profile of validation concerns. For example, flightdeck startle (UC1) weighs most heavily on objective performance and workload, whereas flightdeck planning (UC2) weighs primarily on workload and information security. UC6, which weighs heavily on workload, SA, and information security, weighs low on explainability. This would seem to make intuitive sense, depending on how the prototype is eventually realised. Details of each of the UC validation survey results (including category and objective item results) are shown in Annex D separately for each of the UCs.

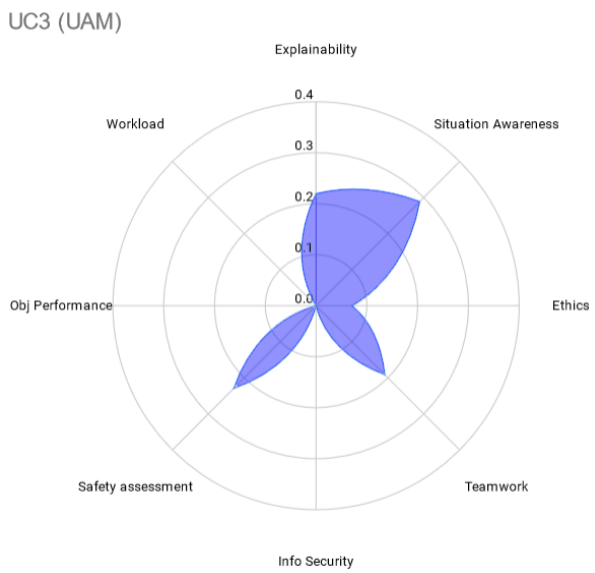
Annex C presents the blank UC survey. Within each of the eight categories, individual HAIT issues are listed (the number of issues ranges from 1 to 13 per category), and corresponding EASA guidelines reference is provided.

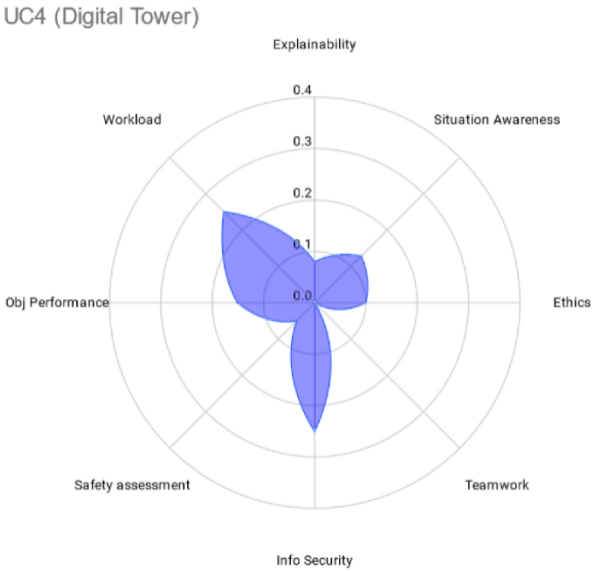
This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Annex D provides detailed survey results, broken out by individual Use Case. Again, rankings for flagged items are abbreviated either H (High) or M (Moderate) in perceived criticality. Additionally, the T3.5 team conducted a post-hoc subjective review of the UC survey responses and compared these to the Use Case descriptions (produced within T6.1 Scenario design for each use case), in which development teams had made an initial assessment of: System goals; Data and time requirements; HAIT specific issues such as user description and system behaviour specs; Interface and communications issues; and an initial description of the use case scenario. Based on this post-hoc review, the T3.5 team recommended some additional items for consideration in individual Use Cases. These recommended items are highlighted in red in Annex D.









This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

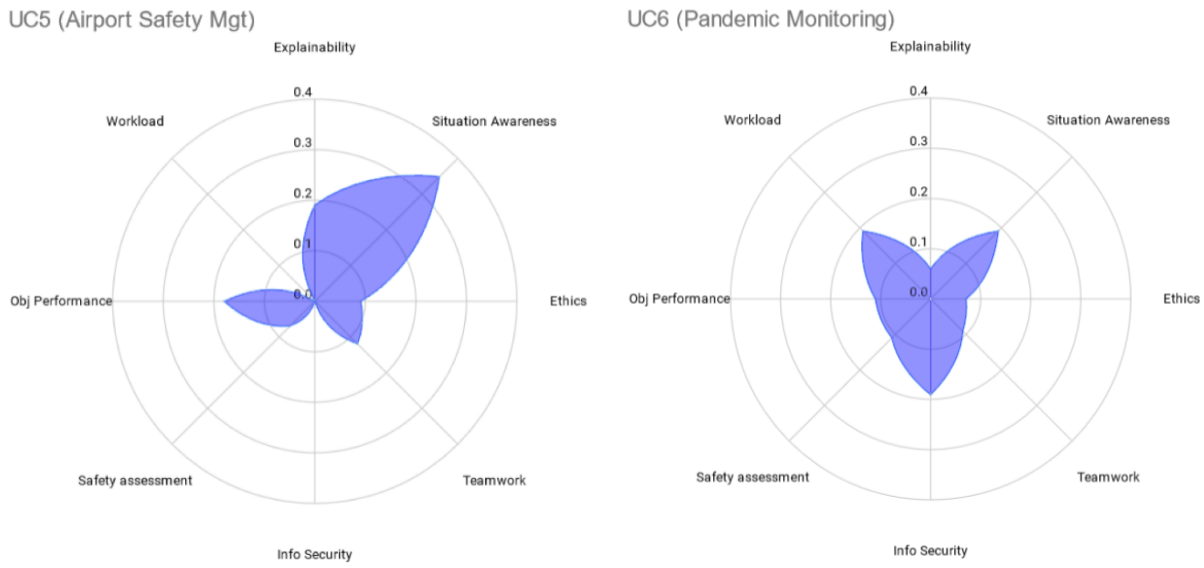


Figure 7a-f: UC validation category profiles, for each UC

Tables 2a-f, below, summarise the high-level survey results from each Use Case. For each, the highest ranked items (High criticality, indicated by 'H' and orange highlighting) are identified within each category. Any additional recommended items are included (indicated by a checkmark and red highlighting).

Category		Guidance
Explainability / Transparency		
	H	Clear and unanmbiguous presentation of explanations
	H	Ensure validity of explanation
	H	Provide timely information on unsafe operating conditions
	☑	Monitor outputs wrt operational performance boundaries and indicate deviations
Situation awareness		
	H	Maintain **operator** Sitiation awareness
Ethics		
	H	Perform ethics-based trustworthiness assessment
	H	Ensure compliance with data protection regulations
	☑	Assess risk of de-skilling
Collaboration and teamwork		
	☑	Ensure AI can build its own Situation Awareness
	☑	Ensure AI can modify its own Situation Awareness
Workload		
	H	Minimise work overload or underload
Information Security		
Safety assessment and assurance		
	H	Identify metrics of AI performance
	H	Identify failure modes and uncertainties
Objective performance criteria		
	H	Maintain reliability
	H	Ensure accuracy

Table 2a. High priority and recommended items, for Use Case 1 (Flightdeck Startle).

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Category		Guidance
Explainability / Transparency		
	H	Define explanations timing according to situation, end user needs, operational impact
	H	Monitor inputs wrt ODD and indicate deviations
	H	Monitor outputs wrt operational performance boundaries and indicate deviations
	☑	Clear and unanmbiguous presentation of explanations
	☑	Ensure validity of explanation
	☑	Provide timely information on unsafe operating conditions
	☑	Provide instructions/training to handle indications of input/output monitoring
	☑	Characterize explainability needs
	☑	Customisation of explanation level of details (if XAI adaptability/adaptiveness is available)
	☑	Enable explanation and details upond user request
	☑	Indicate degree of reliability of explanation
Situation awareness		
	H	Maintain shared situation awareness
Ethics		
	H	Perform ethics-based trustworthiness assessment
	H	Ensure no unfair bias
	☑	Identify new skills
Collaboration and teamwork		
	H	Ensure two-way cross check of proposals
	H	Ensure bi-directional communication
	☑	Identify suboptimal performance or abnormal operation
	☑	Ensure AI can build its own Situation Awareness
	☑	Ensure AI can modify its own Situation Awareness
Workload		
	H	Minimise work overload or underload
Information Security		
	H	Identify and address information security threats introduced by AI usage
	H	Mitigation plan for information security risks introduced by AI usage
Safety assessment and assurance		
	H	Identify failure modes and uncertainties
	H	Specify contingency / mitigation plan for off-nominal data
Objective performance criteria		
	H	Ensure accuracy
	H	Ensure efficiency
	☑	Maintain reliability
	☑	Minimise response time

Table 2b. High priority and recommended items, for Use Case 2 (Flightdeck Planning).

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Category		Guidance
Explainability /		
	H	Demonstrate relevance of explanation for decision/action
	H	Define explanations timing according to situation, end user needs, operational impact
	H	Characterize explainability needs
	H	Define level of abstraction of explanations according to task, situation, trust, expertise of
Situation awareness		
	H	Maintain **operator** Sitiation awareness
Ethics		
	H	Identify new skills
Collaboration and teamwork		
	H	Ensure two-way cross check of proposals
	H	Ensure bi-directional communication
Workload		
Information Security		
Safety assessment and assurance		
	H	Identify failure modes and uncertainties
	H	Specify contingency / mitigation plan for off-nominal data
Objective performance		

Table 2c. High priority and recommended items, for Use Case 3 (UAM).

Category		Guidance
Explainability /		
	H	Characterize explainability needs
	H	Clear and unanmbiguous presentation of explanations
	☑	Monitor outputs wrt operational performance boundaries and indicate deviations
Situation awareness		
	H	Maintain **operator** Sitiation awareness
Ethics		
	H	Perform ethics-based trustworthiness assessment
	H	Ensure no unfair bias
	☑	Identify new skills
	☑	Identify potential health or environmental impacts
Collaboration and teamwork		
		Ensure two-way cross check of proposals
Workload		
	H	Minimise work overload or underload
Information Security		
	H	Identify and address information security threats introduced by AI usage
	H	Mitigation plan for information security risks introduced by AI usage
Safety assessment and assurance		
	H	Identify failure modes and uncertainties
Objective performance		
	H	Maintain reliability
	H	Ensure accuracy

Table 2d. High priority and recommended items, for Use Case 4 (Digital Tower).

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Category	Guidance	
Explainability /		
	H	Demonstrate relevance of explanation for decision/action
	H	Characterize explainability needs
	H	Define level of abstraction of explanations according to task, situation, trust, expertise of
	H	Clear and unanmbiguous presentation of explanations
	H	Ensure validity of explanation
	H	Provide timely information on unsafe operating conditions
	H	Provide instructions/training to handle indications of input/output monitoring
	☑	Enable explanation and details upond user request
	☑	Indicate degree of reliability of explanation
Situation awareness		
	H	Maintain **operator** Sitiation awareness
	H	Maintain **shared** situation awareness
Ethics		
	H	Ensure compliance with data protection regulations
	H	Ensure no unfair bias
Collaboration and teamwork		
	H	Ensure two-way cross check of proposals
	H	Identify suboptimal performance or abnormal operation
Workload		
		Minimise work overload or underload
Information Security		
	☑	Identify and address information security threats introduced by AI usage
	☑	Mitigation plan for information security risks introduced by AI usage
	☑	Verification of security support/mitigation actions
Safety assessment and assurance		
	H	Identify metrics of AI performance
Objective performance		
	H	Ensure accuracy
	H	Ensure efficiency

Table 2e. High priority and recommended items, for Use Case 5 (Airport Safety).

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Category		Guidance
Explainability /		
	H	Demonstrate relevance of explanation for decision/action
	H	Define level of abstraction of explanations according to task, situation, trust, expertise of
	H	Clear and unanmbiguous presentation of explanations
	H	Ensure validity of explanation
Situation awareness		
	H	Maintain **operator** Sitiation awareness
	H	Maintain **shared** situation awareness
Ethics		
	H	Identify new skills
	H	Ensure compliance with data protection regulations
	H	Ensure no unfair bias
	☑	Identify potential health or environmental impacts
Collaboration and teamwork		
	H	Ensure two-way cross check of proposals
	H	Ensure AI can build its own Situation Awareness
	H	Ensure AI can modify its own Situation Awareness
Workload		
	H	Minimise work overload or underload
Information Security		
	H	Identify and address information security threats introduced by AI usage
	H	Mitigation plan for information security risks introduced by AI usage
	H	Verification of security support/mitigation actions
Safety assessment and assurance		
	H	Identify failure modes and uncertainties
	H	Specify contingency / mitigation plan for off-nominal data
	H	Identify metrics of AI performance
Objective performance		
	H	Ensure accuracy
	H	Ensure classification performance
	H	Minimise response time

Table 2f. High priority and recommended items, for Use Case 6 (Pandemic Monitoring).

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

4. Discussion

HAIKU intentionally started from a set of Use Cases that seemed to capture a range of AI levels, and a range of potential HAIT concerns. This was reflected in the UC teams' responses. Two of the UCs (4,5) are aiming for level 1B AI, in which the human retains full authority over decision and action selection. Another two (1,3) are aiming for level 2A AI, in which the human can override these functions. Finally, two UCs (2,6) are aiming for level 2B AI, in which the human has only partial ability to override these functions.

Use Cases also varied in their self-reported identification of critical HAIT issues. As shown in the spider graphs of figures 7a-f, each UC presented a unique HAIT issues profile.

The aim of this task 3.5 was to get a snapshot of the potential HAIT issues associated with each Use Case. Again, this was not meant to be prescriptive, only to help each UC identify the issues that might be most relevant to their case. Given that each Use Case is evolving, the team will consider readministering (perhaps an expanded version of) this survey, as UCs continue development, and focusing more clearly on validation.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

References

Bång, M., Magnus Lundberg, J., Hilburn, B., Humm, E., Mallozzi, G. & Arrigoni, V. (2023). Human-AI Teaming Framework and Design Document. Deliverable 3.1, Human AI Knowledge and Understanding (HAIKU) project.

EASA (2023). EASA Concept Paper: First usable guidance for Level 1 & 2 machine learning applications (proposed Issue 2), released 24 Feb 2023 (https://www.easa.europa.eu/en/newsroom-and-events/news/easa-artificial-intelligence-concept-paper-proposed-issue-2-open). European Union Aviation Safety Agency.

EASA (2020). Artificial Intelligence Roadmap: A Human-centric Approach to AI in Aviation. European Union Aviation Safety Agency. Version 1.0. February 2020.

National Academies of Sciences (2022). Human-AI Teaming: State-of-the-Art and Research Needs. Washington, DC: US National Research Council, National Academies Press. https://doi.org/10.17226/26355.

Parasuraman, R., Sheridan, T.B. & Wickens, C.D. (2000). A Model for Types and Levels of Human Interaction with Automation. In IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Human, vol 30, 3, May 2000.

Sheridan, T.B. & Verplank, W.L. (1978). Human and Computer Control of Undersea Teleoperators. Massachusetts Institute of Technology, Engineering Psychology programs, Office of Naval Research.

Venditti R., Arrigoni V., & Cirillo M.(2023). Concepts of Intelligent Assistants. Deliverable 3.2, Human AI Knowledge and Understanding (HAIKU) project.

Westin, C.A.L. et al. (2013). NALA 3D Validation. Multidimensional Framework for Advanced SESAR Automation (MUFASA) Deliverable 3.2.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Annex A: Survey instructions to UC

HAIKU WP3.5 is developing a provisional framework for validating the Use Case prototypes. This requires identifying the appropriate methods, metrics, and success criteria for validation.

To start this process, the T3.5 team is surveying each of the UCs, to identify the main (and UC-specific) issues to address in validation. These issues come out of the T3.1 state-of-the-art review, in parallel with EASA's guidelines for AI development.

We are obviously not intending a full validation. Instead, think of this as a tryout, in which we identify the most salient / critical / unique issues associated with each UC (hopefully there is some variability across UCs). T3.5 then suggests methods and metrics for each UC to consider.

See the attached Spreadsheet. We have tried to take EASA's relevant guidance (column C), and group it in eight categories (column A). For each of the eight categories, we have listed the related concepts (column G) that came out of the T3.1 review. Notice these concepts in Column G do not map line-for-line with EASA's guidance, but instead fall under the general category.

Here is what we ask you to do:

- Take each category one by one.
- For each category, identify the most potentially critical concepts (column G). Criticality is based on frequency and / or criticality (i.e., the consequence or outcome severity).
- If you can, please rank the top few (max 4 or 5?) concepts. Put a number in the box beside the concept.
- Feel free to append notes to explain why a concept was identified as critical (Frequency? Consequences?)
- Next, look at EASA's guidance in column C, also grouped by category.
- Please also identify and rank a few of the top guidance items that seem most critical to your use case.
- Similarly, feel free to annotate these guidance rankings.
- Finally, notice that most of these items are relevant across AI levels (see column E). The exception is the collaboration and teamwork category, which does not apply to level 1A/1B AI. So you can perhaps skip these items for your Use Case.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Annex B: AI Classification Worksheet

Choose the classification(s) that best capture the AI level of your use case. This scheme is based on EASA (2023), which classifies AI by its process (input, decision, action) and its authority (to decide / act).

Check all that apply

- ☐ Level 1A *Human augmentation* (human has **full** authority)
  - ☐ support to information **acquisition**
  - ☐ support to information **analysis**
- ☐ Level 1B *Human assistance* (human has **full** authority)
  - ☐ Support to **decision** selection
  - ☐ Support to **action** selection
- ☐ Level 2A *Human – AI Cooperation* (human has **full** authority)
  - ☐ Cooperative overridable automatic **decision** selection
  - ☐ Cooperative overridable automatic **action** selection

> ***Level 2a: Human-AI cooperation***: cooperation is a process in which the AI-based system works to help the end user accomplish his or her own goal. The AI-based system works according to a predefined task allocation pattern with informative feedback to the end user on the decisions and/or actions implementation. The cooperation process follows a directive approach. Cooperation does not imply a shared situational awareness between the end user and the AI-based system. Communication is not a paramount capability for cooperation.

- ☐ Level 2B *Human – AI Collaboration* (human has **partial** authority)
  - ☐ Collaborative overridable automatic **decision** selection
  - ☐ Collaborative overridable automatic **action** selection

> ***Level 2B: Human-AI collaboration***: collaboration is a process in which the human end user and the AI-based system work together and jointly to achieve a common goal (or work individually on a defined goal) and solve a problem through co-constructive approach. Collaboration implies the capability to share situational awareness and to readjust strategies and task allocation in real time. Communication is paramount to share valuable information needed to achieve the goal, to share ideas and expectations.

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Annex C: Use Case Survey

Category	Guidance	EASA ref	AI Level	Related concepts
Explainability / Transparency	Characterise explainability needs	EXP-05	1B-2B	Explainability
	Clear and unambiguous presentation of explanations	EXP-06	1B-2B	Transparency
	Demonstrate relevance of explanation for decision/action	EXP-07	1B-2B	Data availability
	Define the level of abstraction of explanations according to task, situation, trust, expertise of user…	EXP-08	1B-2B	Interpretability
	Customisation of explanation level of details (if XAI adaptability/adaptiveness is available)	EXP-09	1B-2B	Observability
	Define explanations timing according to situation, end user needs, operational impact	EXP-10	1B-2B	Predictability
	Enable explanation and details upon user request	EXP-11	1B-2B	Shared goals
	Ensure validity of explanation	EXP-12	1B-2B	Traceability
	Indicate degree of reliability of explanation	EXP-13	1B-2B	Uncertainty

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

	Monitor inputs with respect to ODD and indicate deviations	EXP-14	1A-2B	Understandability
	Monitor outputs with respect to operational performance boundaries and indicate deviations	EXP-15	1A-2B	
	Provide instructions/training to handle indications of input/output monitoring	EXP-16	1A-2B	
	Provide timely information on unsafe operating conditions	EXP-17	1A-2B	
Situation awareness				
	Maintain **operator** Situation Awareness	na	1A-2B	Complacency / vigilance
	Maintain **shared** situation awareness	na	1A-2B	Complexity, task
				Mental model
				Out-of-the-loop
				Shared intent

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Ethics

Perform ethics-based trustworthiness assessment	ET-01	1A-2B
Identify potential health or environmental impacts	ET-02 / ET-06	1A-2B
Identify impact mitigations	ET-07	1A-2B
Ensure no capability of adaptive learning	ET-03	1A-2B
Ensure compliance with data protection regulations	ET-04	1A-2B
Ensure no unfair bias	ET-05	1A-2B
Identify new skills	ET-08	1A-2B
Assess risk of de-skilling	ET-09	1A-2B

- Accountability & auditability
- AI bias
- Data governance
- Data integrity
- Fairness
- Responsibility / liability
- Health impacts
- Environmental impacts
- De-skilling / new skill requirements

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Collaboration and teamwork

Ensure two-way cross check of proposals	HF-04/05/06/11	2A-2B
Identify suboptimal performance or abnormal operation	HF-05 / HF-06	2B
Ensure bi-directional communication	HF-02	2B
Ensure AI can build its own Situation Awareness	HF-01	2B
Ensure AI can modify its own Situation Awareness	HF-03	2B
Notify operator of his / her misunderstanding	HF-12	2B

- Acceptance
- Autonomy
- Complexity, perceived
- Individual differences
- Reliability
- Reliance, over / under
- Self-confidence
- Trust
- Bi-directional communications

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

				Boundary limitations and expectations
Workload	Minimise work overload or underload	na	1A-2B	Workload extremes Vigilance Complacency
Information Security	Identify and address information security threats introduced by AI usage	IS-01	1A-2B	Data integrity
	Mitigation plan for information security risks introduced by AI usage	IS-02	1A-2B	
	Verification of security support/mitigation actions	IS-03	1A-2B	

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Safety assessment and assurance	Identify metrics of AI performance	SA-01-2	1A-2B	
	Specify contingency / mitigation plan for off-nominal data	SA-01-1; SA-01-5	1A-2B	Failure modes
	Identify failure modes and uncertainties	SA-01-4; SA-01-6; SA-01-8	1A-2B	Contingency plans
	Specify data needed for ongoing safety assessment	ICSA-01	1A-2B	
	Define Safety assessment methodology (target values, threshold, evaluation periods, etc)	ICSA-02	1A-2B	
Objective performance criteria		SA-01-2		
	Ensure accuracy	na	1A-2B	System accuracy
	Ensure classification performance	na	1A-2B	System classification performance

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

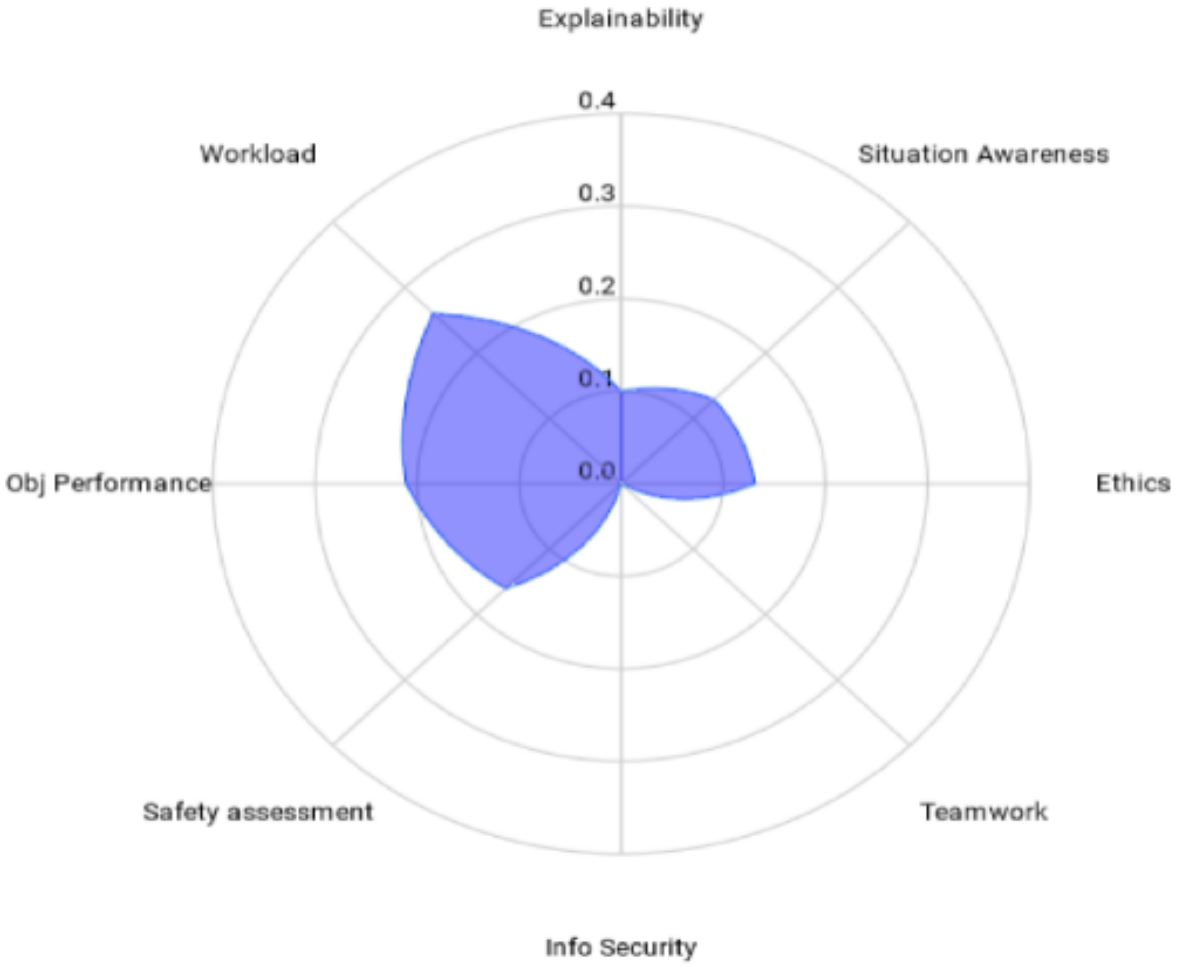
Ensure efficiency	na	1A-2B	System efficiency
Maintain reliability	na	1A-2B	System reliability
Minimise response time	na	1A-2B	System latency (response time)

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Annex D: Survey results, by Use Case

Note: proposed additional elements in red

Use Case 1: Flightdeck startle

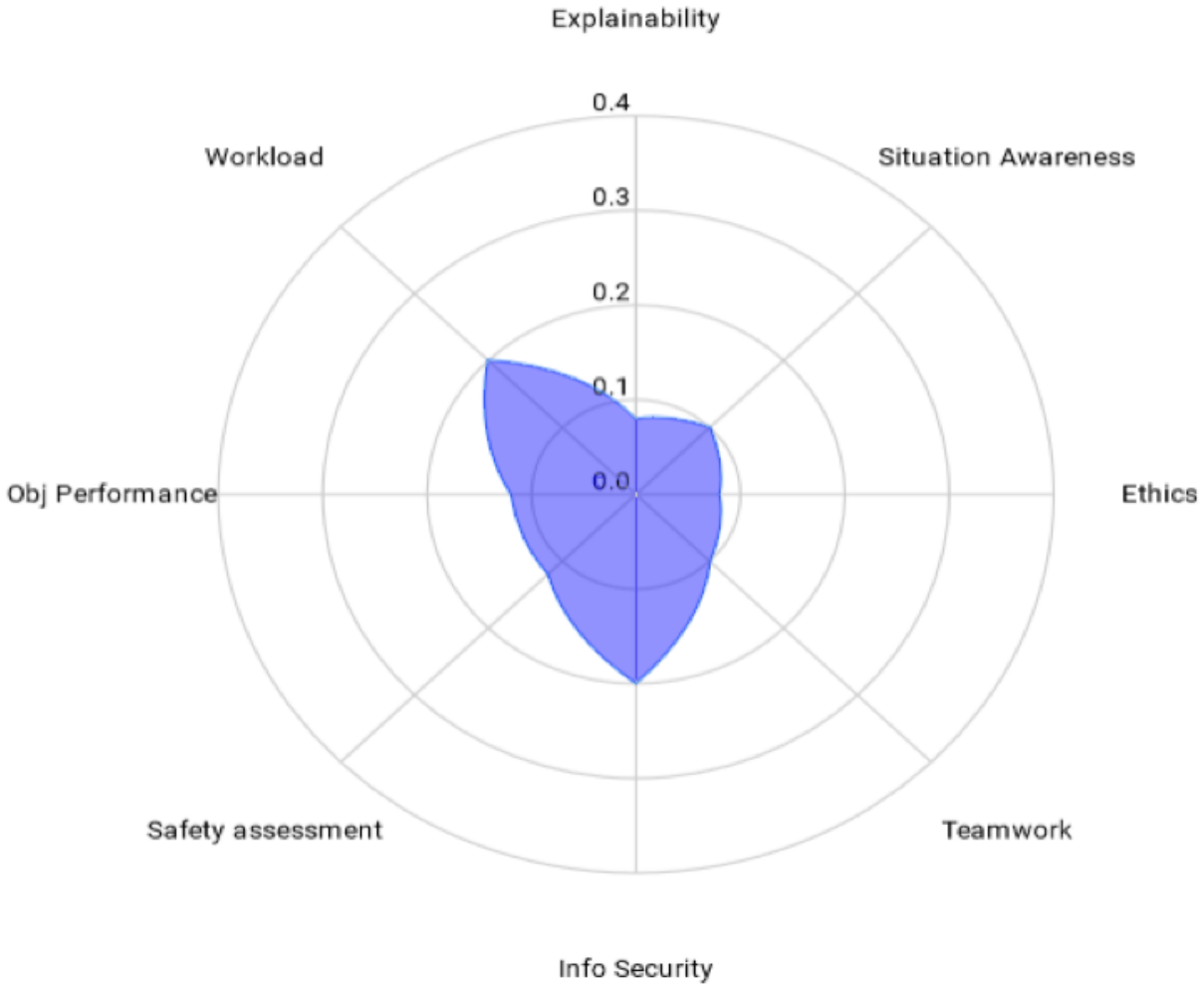


This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Category		Guidance		Related concepts
Explainability / Transparency	H	Clear and unanmbiguous presentation of explanations	H	Understandability
	H	Ensure validity of explanation	H	Interpretability
	H	Provide timely information on unsafe operating conditions	M	Explainability
	M	Demonstrate relevance of explanation for decision/action	M	Data availability
	M	Define explanations timing according to situation, end user needs, operational impact	▯	Transparency
	M	Provide instructions/training to handle indications of input/output monitoring	▯	Observability
		Characterize explainability needs	▯	Predictability
		Define level of abstraction of explanations according to task, situation, trust, expertise of user...		Shared goals
		Customisation of explanation level of details (if XAI adaptability/adaptiveness is available)	▯	Traceability
		Enable explanation and details upond user request		Uncertainty
		Indicate degree of reliability of explanation		
		Monitor inputs wrt ODD and indicate deviations		
	▯	Monitor outputs wrt operational performance boundaries and indicate deviations		
Situation awareness	H	Maintain **operator** Sitiation awareness	H	Complacency / vigilance
		Maintain **shared** situation awareness	M	Complexity, task
			▯	Mental model
			▯	Out-of-the-loop
				Shared intent
Ethics	H	Perform ethics-based trustworthiness assessment	H	Data governance
	H	Ensure compliance with data protection regulations	H	Responsibility / liability
	M	Identify potential health or environmental impacts	M	Health impacts
	M	Ensure no unfair bias		Data integrity
		Identify impact mitigations		Fairness
		Ensure no capability of adaptive learning		Accountability & auditability
		Identify new skills		Environmental impacts
	▯	Assess risk of de-skilling		AI bias
Collaboration and teamwork		Ensure two-way cross check of proposals	▯	Acceptance
		Identify suboptimal performance or abnormal operation		Autonomy
		Ensure bi-directional communication		Complexity, perceived
	▯	Ensure AI can build its own Situation Awareness		Individual differences
	▯	Ensure AI can modify its own Situation Awareness		Reliability
		Notify operator of his / her misunderstanding	▯	Reliance, over / under
				Self-confidence
			▯	Trust
				Bi-directional communications
			▯	Boundary limitations and expectations
Workload	H	Minimise work overload or underload	H	Workload extremes
				Vigilance
				Complexity
Information Security		Identify and address information security threats introduced by AI usage		Data integrity
		Mitigation plan for information security risks introduced by AI usage		
		Verification of security support/mitigation actions		
Safety assessment and assurance	H	Identify metrics of AI performance	H	Failure modes
	H	Identify failure modes and uncertainties	M	Contingency plans
	M	Specify contingency / mitigation plan for off-nominal data		
Objective performance	H	Maintain reliability	H	System accuracy
	H	Ensure accuracy	H	System reliability
	M	Ensure classification performance	M	System classification performance
	M	Minimise response time	M	System latency (response time)
		Ensure efficiency		System efficiency

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Use Case 2: Flightdeck planning



This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Category		Guidance		Related concepts
Explainability / Transparency	H	Define explanations timing according to situation, end user needs, operational impact	H	Predictability
	H	Monitor inputs wrt ODD and indicate deviations	H	Shared goals
	H	Monitor outputs wrt operational performance boundaries and indicate deviations	M	Explainability
	M	Demonstrate relevance of explanation for decision/action	D	Understandability
	M	Define level of abstraction of explanations according to task, situation, trust, expertise of user...	D	Interpretability
	D	Clear and unanmbiguous presentation of explanations		Data availability
	D	Ensure validity of explanation	D	Transparency
	D	Provide timely information on unsafe operating conditions	D	Observability
	D	Provide instructions/training to handle indications of input/output monitoring		Traceability
	D	Characterize explainability needs	D	Uncertainty
	D	Customisation of explanation level of details (if XAI adaptability/adaptiveness is available)		
	D	Enable explanation and details upond user request		
	D	Indicate degree of reliability of explanation		
Situation awareness	H	Maintain **shared** situation awareness	H	Mental model
		Maintain **operator** Sitiation awareness	H	Shared intent
			M	Out-of-the-loop
				Complexity, task
				Complacency / vigilance
Ethics	H	Perform ethics-based trustworthiness assessment	H	Accountability & auditability
	H	Ensure no unfair bias	M	AI bias
	M	Assess risk of de-skilling		Data governance
		Ensure compliance with data protection regulations		Responsibility / liability
		Identify potential health or environmental impacts		Health impacts
		Identify impact mitigations		Data integrity
		Ensure no capability of adaptive learning		Fairness
	D	Identify new skills		Environmental impacts
Collaboration and teamwork	H	Ensure two-way cross check of proposals	H	Bi-directional communications
	H	Ensure bi-directional communication	H	Boundary limitations and expectations
	M	Notify operator of his / her misunderstanding	M	Reliance, over / under
	D	Identify suboptimal performance or abnormal operation	D	Acceptance
	D	Ensure AI can build its own Situation Awareness		Autonomy
	D	Ensure AI can modify its own Situation Awareness		Complexity, perceived
			D	Individual differences
			D	Reliability
				Self-confidence
			D	Trust
Workload	H	Minimise work overload or underload	H	Workload extremes
			H	Vigilance
			M	Complexity
Information Security	H	Identify and address information security threats introduced by AI usage		Data integrity
	H	Mitigation plan for information security risks introduced by AI usage		
	M	Verification of security support/mitigation actions		
Safety assessment and assurance	H	Identify failure modes and uncertainties	H	Failure modes
	H	Specify contingency / mitigation plan for off-nominal data	M	Contingency plans
	M	Identify metrics of AI performance		
		Specify data needed for ongoing safety assessment		
		Define Safety assessment methodology (target values, threshold, evaluation periods, etc)		
Objective performance	H	Ensure accuracy	H	System accuracy
	H	Ensure efficiency	H	System efficiency
	M	Ensure classification performance	M	System classification performance
	D	Maintain reliability	D	System reliability
	D	Minimise response time	D	System latency (response time)

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

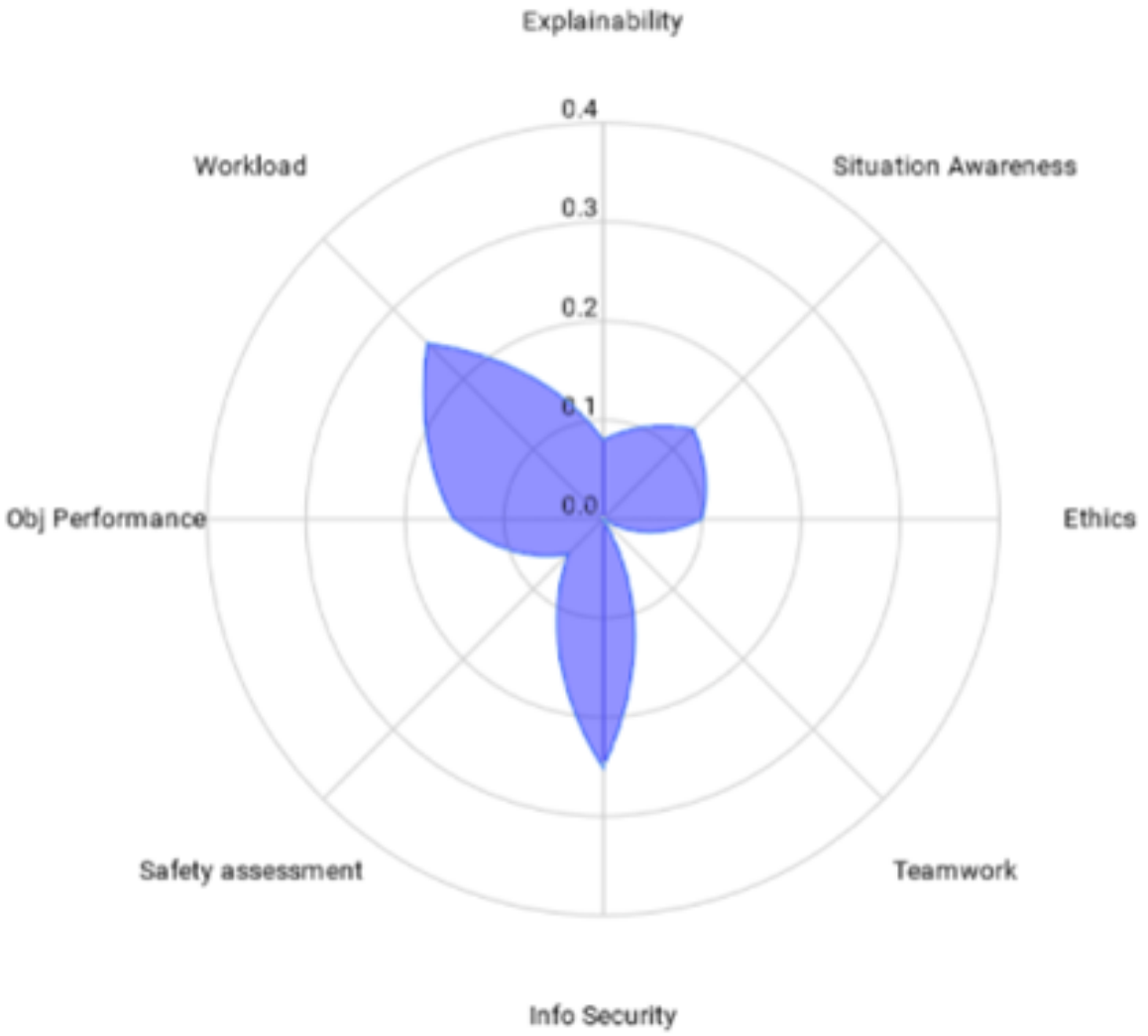
Use Case 3: Urban Air Mobility (UAM)

Note: evolving ConOps - No additional guidelines currently suggested

Category	Guidance	Related concepts
Explainability /		
	H Demonstrate relevance of explanation for decision/action	H Understandability
	H Define explanations timing according to situation, end user needs, operational impact	H Explainability
	H Characterize explainability needs	H Shared goals
	H Define level of abstraction of explanations according to task, situation, trust, expertise	H Traceability
	Clear and unanmbiguous presentation of explanations	Interpretability
	Ensure validity of explanation	Data availability
	Provide timely information on unsafe operating conditions	Transparency
	Provide instructions/training to handle indications of input/output monitoring	Observability
	Customisation of explanation level of details (if XAI adaptability/adaptiveness is	Predictability
	Enable explanation and details upond user request	Uncertainty
	Indicate degree of reliability of explanation	
	Monitor inputs wrt ODD and indicate deviations	
	Monitor outputs wrt operational performance boundaries and indicate deviations	
Situation awareness		
	H Maintain **operator** Sitiation awareness	H Out-of-the-loop
	Maintain **shared** situation awareness	H Shared intent
		Complacency / vigilance
		Complexity. task
		Mental model
Ethics		
	H Identify new skills	Data governance
	Perform ethics-based trustworthiness assessment	Responsibility / liability
	Ensure compliance with data protection regulations	Health impacts
	Identify potential health or environmental impacts	Data integrity
	Ensure no unfair bias	Fairness
	Identify impact mitigations	Accountability & auditability
	Ensure no capability of adaptive learning	Environmental impacts
	Assess risk of de-skilling	AI bias
Collaboration and		
	H Ensure two-way cross check of proposals	H Bi-directional communications
	Identify suboptimal performance or abnormal operation	Acceptance
	H Ensure bi-directional communication	Autonomy
	Ensure AI can build its own Situation Awareness	Complexity. perceived
	Ensure AI can modify its own Situation Awareness	Individual differences
	Notify operator of his / her misunderstanding	Reliability
		Reliance, over / under
		Self-confidence
		Trust
		Boundary limitations and expectations
Workload		
	Minimise work overload or underload	Workload extremes
		Vigilance
		Complexity
Information Security		
	Identify and address information security threats introduced by AI usage	Data integrity
	Mitigation plan for information security risks introduced by AI usage	
	Verification of security support/mitigation actions	
Safety assessment and assurance		
	H Identify failure modes and uncertainties	H Failure modes
	H Specify contingency / mitigation plan for off-nominal data	Contingency plans
	Identify metrics of AI performance	
	Specify data needed for ongoing safety assessment	
	Define Safety assessment methodology (target values, threshold, evaluation periods, etc)	
Objective performance		
	Maintain reliability	System accuracy
	Ensure accuracy	System reliability
	Ensure classification performance	System classification performance
	Minimise response time	System latency (response time)
	Ensure efficiency	System efficiency

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Use Case 4: Digital Tower

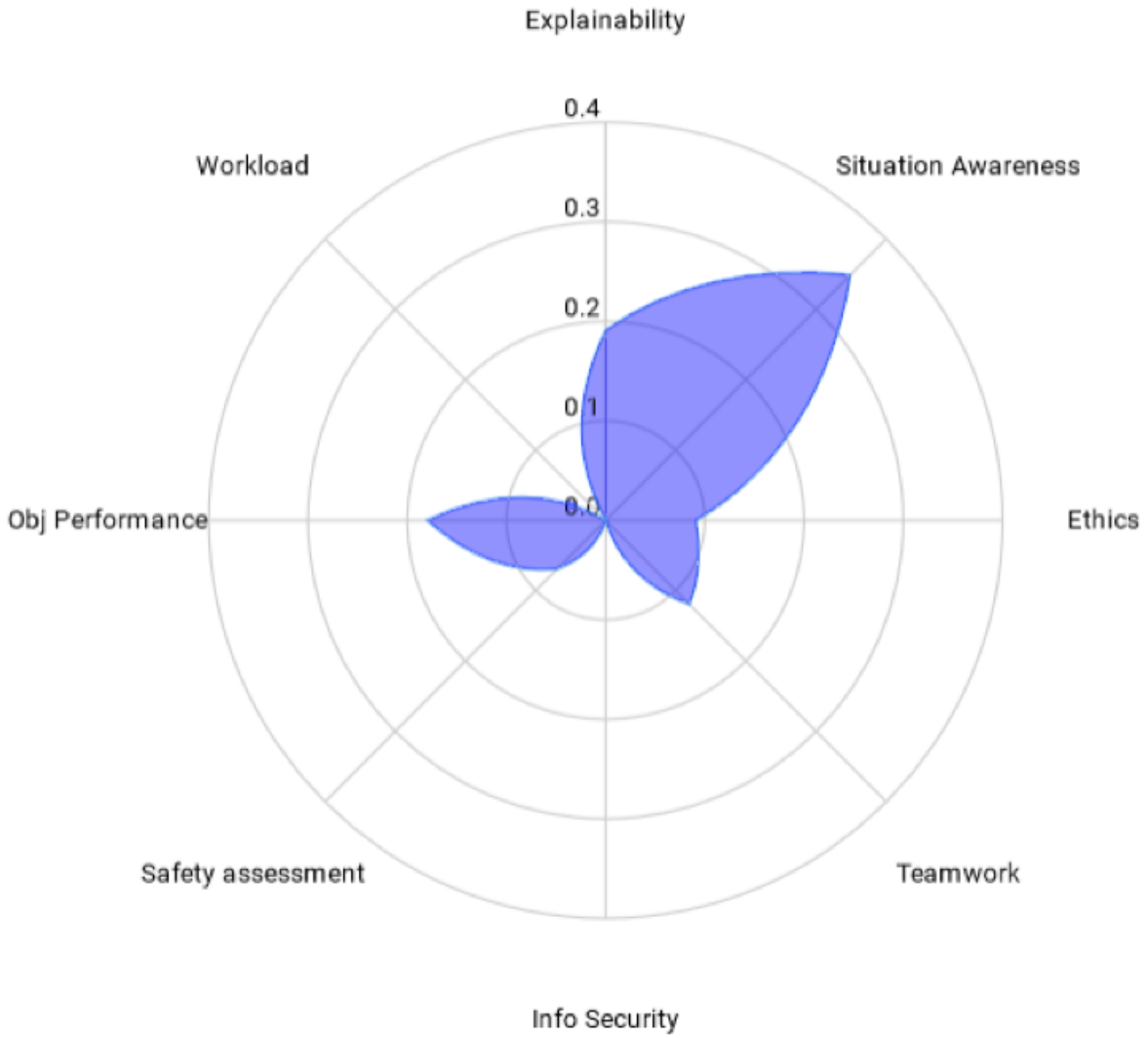


This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Category		Guidance		Related concepts
Explainability / Interpretability / Observability	H	Characterize explainability needs	H	Explainability
	H	Clear and unambiguous presentation of explanations	H	Traceability
	M	Define explanations timing according to situation, end user needs, operational impact	M	Understandability
	M	Provide timely information on unsafe operating conditions	M	Predictability
		Demonstrate relevance of explanation for decision/action		Shared goals
		Define level of abstraction of explanations according to task, situation, trust, expertise		Interpretability
		Ensure validity of explanation		Data availability
		Provide instructions/training to handle indications of input/output monitoring		Transparency
		Customisation of explanation level of details (if XAI adaptability/adaptiveness is		Observability
		Enable explanation and details upond user request		Uncertainty
		Indicate degree of reliability of explanation		
		Monitor inputs wrt ODD and indicate deviations		
	D	Monitor outputs wrt operational performance boundaries and indicate deviations		
Situation awareness	H	Maintain **operator** Sitiation awareness	H	Complacency / vigilance
		Maintain **shared** situation awareness		Out-of-the-loop
				Shared intent
				Complexity, task
				Mental model
Ethics	H	Perform ethics-based trustworthiness assessment	H	Responsibility / liability
	H	Ensure no unfair bias	H	Accountability & auditability
	M	Ensure compliance with data protection regulations	M	Data governance
	D	Identify new skills		Health impacts
	D	Identify potential health or environmental impacts		Data integrity
	D	Identify impact mitigations		Fairness
		Ensure no capability of adaptive learning		Environmental impacts
		Assess risk of de-skilling		AI bias
Collaboration and teamwork		Ensure two-way cross check of proposals		Bi-directional communications
		Identify suboptimal performance or abnormal operation		Acceptance
		Ensure bi-directional communication		Autonomy
		Ensure AI can build its own Situation Awareness		Complexity, perceived
		Ensure AI can modify its own Situation Awareness		Individual differences
		Notify operator of his / her misunderstanding		Reliability
				Reliance, over / under
				Self-confidence
				Trust
				Boundary limitations and expectations
Workload	H	Minimise work overload or underload	H	Vigilance
			H	Complexity
			M	Workload extremes
Information Security	H	Identify and address information security threats introduced by AI usage	H	Data integrity
	H	Mitigation plan for information security risks introduced by AI usage		
	M	Verification of security support/mitigation actions		
Safety assessment and assurance	H	Identify failure modes and uncertainties	H	Failure modes
		Specify contingency / mitigation plan for off-nominal data		Contingency plans
		Identify metrics of AI performance		
		Specify data needed for ongoing safety assessment		
		Define Safety assessment methodology (target values, threshold, evaluation periods, etc)		
Objective performance	H	Maintain reliability	H	System accuracy
	H	Ensure accuracy	H	System reliability
		Ensure classification performance	M	System classification performance
		Minimise response time		System latency (response time)
	M	Ensure efficiency		System efficiency

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Use Case 5: Airport Safety Management

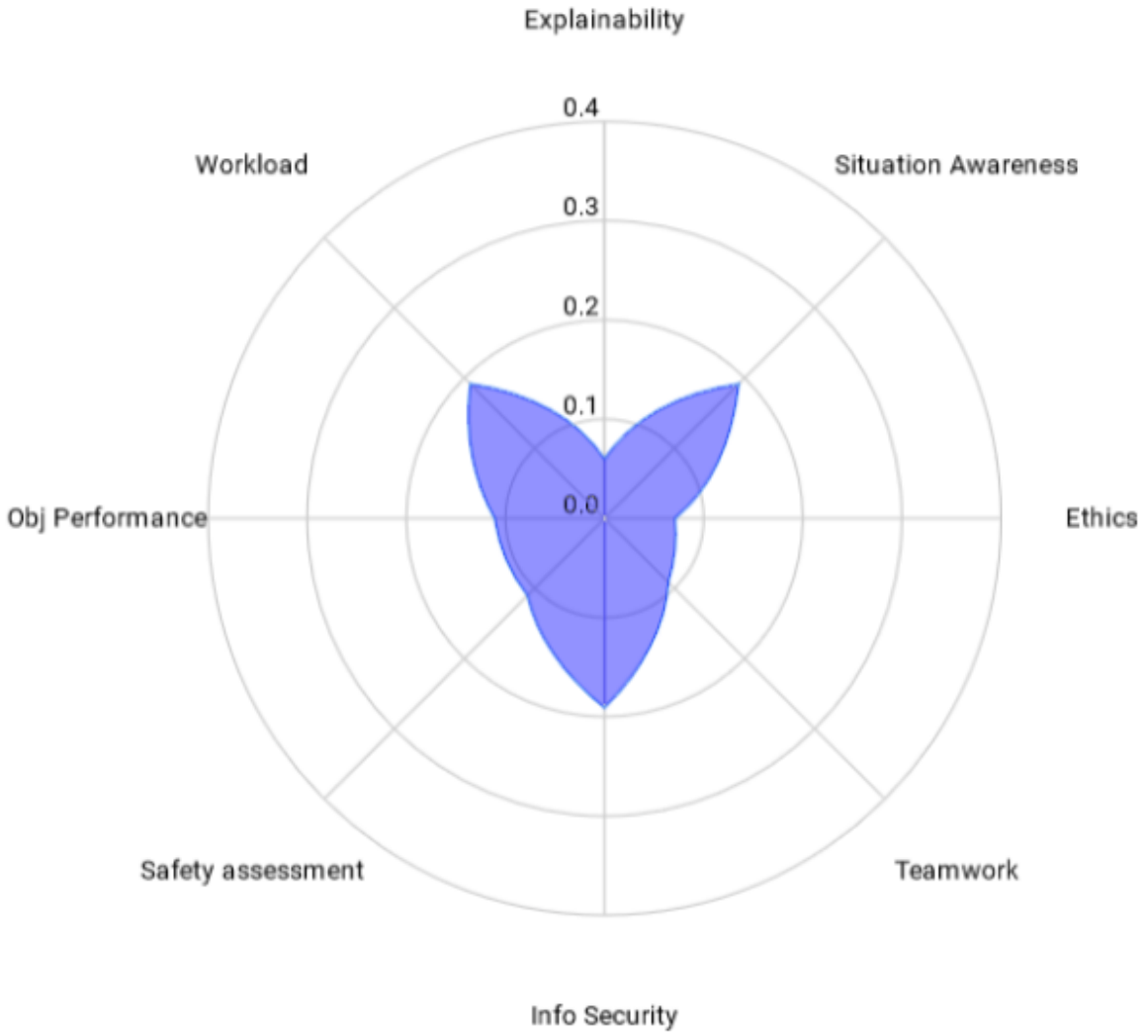


This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Category		Guidance		Related concepts
Explainability /				
	H	Demonstrate relevance of explanation for decision/action	H	Understandability
	H	Characterize explainability needs	H	Interpretability
	H	Define level of abstraction of explanations according to task, situation, trust, expertise	H	Predictability
	H	Clear and unanmbiguous presentation of explanations	H	Uncertainty
	H	Ensure validity of explanation		Explainability
	H	Provide timely information on unsafe operating conditions		Shared goals
	H	Provide instructions/training to handle indications of input/output monitoring		Traceability
		Define explanations timing according to situation, end user needs, operational impact		Data availability
		Customisation of explanation level of details (if XAI adaptability/adaptiveness is		Transparency
	D	Enable explanation and details upond user request		Observability
	D	Indicate degree of reliability of explanation		
		Monitor inputs wrt ODD and indicate deviations		
		Monitor outputs wrt operational performance boundaries and indicate deviations		
Situation awareness				
	H	Maintain **operator** Sitiation awareness	H	Mental model
	H	Maintain **shared** situation awareness		Out-of-the-loop
				Shared intent
				Complacency / vigilance
				Complexity, task
Ethics				
	H	Ensure compliance with data protection regulations	H	Responsibility / liability
	H	Ensure no unfair bias	H	AI bias
		Identify new skills		Data governance
		Perform ethics-based trustworthiness assessment		Health impacts
		Identify potential health or environmental impacts		Data integrity
		Identify impact mitigations		Fairness
		Ensure no capability of adaptive learning		Accountability & auditability
		Assess risk of de-skilling		Environmental impacts
Collaboration and				
	H	Ensure two-way cross check of proposals	H	Acceptance
	H	Identify suboptimal performance or abnormal operation	H	Trust
		Ensure bi-directional communication		Bi-directional communications
		Ensure AI can build its own Situation Awareness		Autonomy
		Ensure AI can modify its own Situation Awareness		Complexity, perceived
		Notify operator of his / her misunderstanding		Individual differences
				Reliability
				Reliance, over / under
				Self-confidence
				Boundary limitations and expectations
Workload				
		Minimise work overload or underload		Workload extremes
				Vigilance
				Complexity
Information Security				
	D	Identify and address information security threats introduced by AI usage		Data integrity
	D	Mitigation plan for information security risks introduced by AI usage		
	D	Verification of security support/mitigation actions		
Safety assessment and assurance				
	H	Identify metrics of AI performance		Failure modes
		Identify failure modes and uncertainties		Contingency plans
		Specify contingency / mitigation plan for off-nominal data		
		Specify data needed for ongoing safety assessment		
		Define Safety assessment methodology (target values, threshold, evaluation periods, etc)		
Objective performance				
	H	Ensure accuracy		System accuracy
	H	Ensure efficiency		System reliability
		Maintain reliability		System classification performance
		Ensure classification performance		System latency (response time)
		Minimise response time		System efficiency

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Use Case 6: Pandemic monitoring



This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Category	Guidance	Related concepts
Explainability / Interpretability	H Demonstrate relevance of explanation for decision/action H Define level of abstraction of explanations according to task, situation, trust, expertise H Clear and unambiguous presentation of explanations H Ensure validity of explanation Define explanations timing according to situation, end user needs, operational impact Characterize explainability needs Provide timely information on unsafe operating conditions Provide instructions/training to handle indications of input/output monitoring Customisation of explanation level of details (if XAI adaptability/adaptiveness is Enable explanation and details upond user request Indicate degree of reliability of explanation Monitor inputs wrt ODD and indicate deviations Monitor outputs wrt operational performance boundaries and indicate deviations	H Explainability H Interpretability H Data availability H Uncertainty H Predictability Understandability Shared goals Traceability Transparency Observability
Situation awareness	H Maintain **operator** Sitiation awareness H Maintain **shared** situation awareness	H Shared intent H Complacency / vigilance H Complexity, task H Mental model Out-of-the-loop
Ethics	H Identify new skills H Ensure compliance with data protection regulations H Ensure no unfair bias Identify impact mitigations Perform ethics-based trustworthiness assessment I Identify potential health or environmental impacts Ensure no capability of adaptive learning Assess risk of de-skilling	H Health impacts H Data integrity H Fairness H AI bias Data governance Responsibility / liability Accountability & auditability Environmental impacts
Collaboration and teamwork	H Ensure two-way cross check of proposals H Ensure AI can build its own Situation Awareness H Ensure AI can modify its own Situation Awareness Identify suboptimal performance or abnormal operation Ensure bi-directional communication Notify operator of his / her misunderstanding	H Bi-directional communications H Acceptance H Autonomy H Reliability H Self-confidence H Trust Complexity, perceived Individual differences Reliance, over / under Boundary limitations and expectations
Workload	H Minimise work overload or underload	H Workload extremes H Vigilance H Complexity
Information Security	H Identify and address information security threats introduced by AI usage H Mitigation plan for information security risks introduced by AI usage H Verification of security support/mitigation actions	H Data integrity
Safety assessment and assurance	H Identify failure modes and uncertainties H Specify contingency / mitigation plan for off-nominal data H Identify metrics of AI performance Specify data needed for ongoing safety assessment Define Safety assessment methodology (target values, threshold, evaluation periods, etc)	H Failure modes H Contingency plans
Objective performance	H Ensure accuracy H Ensure classification performance H Minimise response time Maintain reliability Ensure efficiency	H System accuracy H System reliability H System classification performance H System latency (response time) H System efficiency

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Annex E: HAIT constructs, and preliminary mapping to assessment methods (cf Bång et al, 2023)e

	Self-report	Query	Checklist	Observation	Behaviour	System performance	Analytic	Physiological
Acceptance	X			X	X	X		
Accessibility			X	X			X	
Accountability			X				X	
Accuracy				X	X	X		
Adversarial training techniques								
Agent capability mismatch								
AI bias			X				X	
Auditability			X				X	
Authority								

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Automation acceptance	X			X	X	X		
Automation reliability								
Automation use								
Automation visibility	X	X	X	X	X			
Automations reliability						X		
Autonomy								
Bias against automation	X	X						X
Bias in AI decision-making			X				X	
Brittleness	X		X	X		X	X	
Calibrated trust	X		X	X	X			X
Certification			X					
Classification performance				X		X	X	
Cognitive processes	X		X	X	X		X	X

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Collaboration	X		X	X	X			
Communicate goals			X	X				
Communication				X				
Complacency and over/under reliance			X	X			X	X
Complexity, perceived	X							
Complexity, task				X		X	X	
Comprehensibility	X	X	X	X	X			
Comprehension	X		X	X	X			
Confidence in AI performance	X		X	X	X			
Confidence in operator manual ability	X		X	X	X			
Consistency				X	X			
Cooperation				X	X			
Coordination	X		X	X	X			

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Costs								
Data availability				X			X	
Data governance			X				X	
Data integrity			X				X	
Decision making effectiveness				X	X		X	
Decision-making biases, AI			X				X	
Dereferral procedures				X	X			
Dispositional trust	X							
Engagement	X			X	X			X
Environmental constraints								
Error trapping and handling	X			X	X	X		
Errors	X			X	X	X	X	

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Ethics			X				X	
Explainability	X	X	X	X	X			
Failure mode model							X	
Fairness			X				X	
Fault tolerance						X	X	
Function allocation							X	
Goal compatibility	X	X	X	X				
Human error probabilities							X	
Human-AI interaction methods	X	X		X			X	
Individual differences	X	X		X	X			
Intelligibility	X	X	X	X	X			
Intent	X	X		X	X			
Interfaces								

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Interpretability	X	X	X	X	X			
Intervention supporting								
Job satisfaction	X	X						
Joint human-AI system performance					X			
Mental model	X	X		X	X			
Misuse (overuse) or disuse	X	X	X	X	X	X		
Mode salience								
Monitoring								
Mutual task coordination								
Mutual trust								
Objective performance								
Observability	X	X	X	X				
Operator experience	X	X						

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Out-of-the-loop	X	X		X	X			
Pedigree	X	X		X	X	X		
Performance					X	X		
Physical coherency				X				
Planned actions	X	X						
Predictability of future actions	X	X	X	X				
Predicted consequences	X	X						
Purpose and goals	X							
Pursuit of shared goals	X	X						
Reliability						X		
Reliance, over / under						X		
Response time						X		
Responsibility			X				X	

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Return-to-manual control	X	X		X	X	X		
Robustness								
Safety					X	X		
Security								
Shared intent	X	X						
Shared situation awareness	X	X	X	X				
Shared understanding	X	X	X	X	X			
Situation awareness	X	X	X	X				X
Skill retention	X	X	X					
Social justice maintenance			X				X	
Strategies								
Strategy mismatch / non-conformance	X	X		X	X			

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Subjective metrics								
System efficiency,								
Task uncertainty	X	X	X					
Team biases								
Team cognitive coherence								
Team decision making								
Team performance								X
Team situation awareness	X	X	X	X				
Team training								
Team trust	X	X	X					X
Teambuilding								
Teamwork processes	X	X	X					X
Test intrusiveness								

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Test reliability								
Time								
Time pressure	X	X	X	X	X			
Traceability								
Trade-offs								
Training, new training requirements								
Training, perturbation training								
Transparency	X	X	X	X				
Trust	X	X		X	X			
Trust, calibrated	X	X		X	X			
Trust, dispositional	X	X						X
Trustworthiness			X				X	
Uncertainty	X	X						

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Understandability	X	X	X	X	X			
Unexpected automation transitions	X	X	X	X	X			
Workload	X	X		X	X		X	X
Workload extremes	X	X		X	X		X	X
Workload management	X	X		X	X		X	X

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332